TASK 1:

pwd: displays the current directory.

$ pwd

/Users/gift

$

vi : Starts the vi editor for editing a file

vi Testfile.txt

touch: touch file sets the current time stamp for a file. Can be used to create an empty file.

Examples below:

- Create a blank file

  $ touch test.txt

  ls

  $ ls -l test.txt

  -rw-r--r-- 1 gift staff 0 Mar 18 12:27 test.txt

- Change access time

  $ touch -c test.txt

  ls -l test.txt

  -rw-r--r-- 1 gift staff 0 Mar 18 12:32 test.txt

- Create lots and lots of files

  $ touch {A..Z}

  $ touch {1..20}

  $ touch {1..1000}.txt

  $ ls

1	11	15	19	3	5	7	9	C	G	K	O	S	W
1.txt	12	16	2	3.txt	5.txt	7.txt	9.txt	D	H	L	P	T	X
10	13	17	2.txt	4	6	8	A	E	I	M	Q	U	Y
10.txt	14	18	20	4.txt	6.txt	8.txt	B	F	J	N	R	V	Z

mkdir : Create a new directory

  $ mkdir gc_dir

  $ ls -l

```
        drwxr-xr-x  2 gift  staff  64 Mar 18 12:46 gc_dir
```

rm : deletes a file or directory

```
$ ls -l
drwxr-xr-x  2 gift  staff  64 Mar 18 12:46 gc_dir
-rw-r--r--  1 gift  staff   0 Mar 18 12:49 testfile.txt
$ rm testfile.txt
$ rm -r gc_dir/
$ ls -l
$
```

ls : displays the contents of a directory

```
$ ls -l
drwxr-xr-x  2 gift  staff  64 Mar 18 12:46 gc_dir
 -rw-r--r--  1 gift  staff   0 Mar 18 12:49 testfile.txt
```

echo : display a line of text/string on standard output or a file.

```
$ echo "display a line of text/string on standard output or a file."
display a line of text/string on standard output or a file.
```

cat : displays the content of a file on the standard output channel (usually the screen).

```
$ cat testfile.txt
This is a test file
re
fgtsss
ddd
$
```

who: display the users that are currently logged into your Unix computer system.

```
$ who
gift     console  Mar 17 21:57
gift     ttys000  Mar 18 12:17
$
```

cd : Changes the working directory

$ pwd

/Users/gift

$cd /Users/gift/work

$pwd]

/Users/gift/work

date: displays the current system date and time

 $ date

Mon Mar 18 14:47:54 CAT 2019

cal: he calendar of a specific month or a whole year.

$ cal

   March 2019

Su Mo Tu We Th Fr Sa

         1 2

 3 4 5 6 7 8 9

10 11 12 13 14 15 16

17 18 19 20 21 22 23

24 25 26 27 28 29 30

31

$ cal 05 2006

mv : moves file/directory to another folder

$ pwd

/Users/gift

ls

testfile.txt
ls work/
mv testfile.txt work/
ls work/
testfile.txt

cp: copies file/directory

ls

testfile.txt

cp testfile.txt testfile2.txt
ls
testfile.txt testfile2.txt

which: searches the path of executable in system paths set in $PATH environment variable.

$ which cat

/bin/cat

$ which vi echo ls

/usr/bin/vi

/bin/echo

/bin/ls

TASK 2:

Word Median

```
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples-2.6.0-cdh5.13.0.jar wordmedian  wordmedian/input
wordmedian/output
19/03/27 06:58:14 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/03/27 06:58:15 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
execute your application with ToolRunner to remedy this.
19/03/27 06:58:15 INFO input.FileInputFormat: Total input paths to process : 1
19/03/27 06:58:16 INFO mapreduce.JobSubmitter: number of splits:1
19/03/27 06:58:16 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1553694939276_0001
19/03/27 06:58:17 INFO impl.YarnClientImpl: Submitted application application_1553694939276_0001
19/03/27 06:58:17 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1553694939276_0001/
19/03/27 06:58:17 INFO mapreduce.Job: Running job: job_1553694939276_0001
19/03/27 06:58:28 INFO mapreduce.Job: Job job_1553694939276_0001 running in uber mode : false
19/03/27 06:58:28 INFO mapreduce.Job:  map 0% reduce 0%
19/03/27 06:58:36 INFO mapreduce.Job:  map 100% reduce 0%
19/03/27 06:58:42 INFO mapreduce.Job:  map 100% reduce 100%
19/03/27 06:58:42 INFO mapreduce.Job: Job job_1553694939276_0001 completed successfully
19/03/27 06:58:42 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=766
                FILE: Number of bytes written=288509
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=30680
                HDFS: Number of bytes written=414
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=4315
                Total time spent by all reduces in occupied slots (ms)=3355
                Total time spent by all map tasks (ms)=4315
                Total time spent by all reduce tasks (ms)=3355
                Total vcore-milliseconds taken by all map tasks=4315
                Total vcore-milliseconds taken by all reduce tasks=3355
                Total megabyte-milliseconds taken by all map tasks=4418560
                Total megabyte-milliseconds taken by all reduce tasks=3435520
        Map-Reduce Framework
                Map input records=223
                Map output records=2278
                Map output bytes=18224
                Map output materialized bytes=766
                Input split bytes=140
                Combine input records=2278
                Combine output records=76
                Reduce input groups=76
                Reduce shuffle bytes=766
                Reduce input records=76
                Reduce output records=76
                Spilled Records=152
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=435
                CPU time spent (ms)=2420
                Physical memory (bytes) snapshot=642392064
                Virtual memory (bytes) snapshot=3154923520
                Total committed heap usage (bytes)=546832384
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=30540
        File Output Format Counters
                Bytes Written=414
The median is: 5
[cloudera@quickstart ~]$ hadoop fs -ls wordmedian/input wordmedian/output
Found 1 items
-rw-r--r--   1 cloudera cloudera      30540 2019-03-27 00:25 wordmedian/input/testfile.txt
Found 2 items
-rw-r--r--   1 cloudera cloudera          0 2019-03-27 06:58 wordmedian/output/_SUCCESS
-rw-r--r--   1 cloudera cloudera        414 2019-03-27 06:58 wordmedian/output/part-r-00000
[cloudera@quickstart
```

WordStandardDeviation

```
[cloudera@quickstart ~]$ hadoop fs -put testfile.txt  /user/cloudera/wordstandarddeviation/input
[cloudera@quickstart ~]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples-2.6.0-cdh5.13.0.jar wordstandarddeviation
wordstandarddeviation/input wordstandarddeviation/output
19/03/18 09:21:39 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/03/18 09:21:40 INFO input.FileInputFormat: Total input paths to process : 1
19/03/18 09:21:40 INFO mapreduce.JobSubmitter: number of splits:1
19/03/18 09:21:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1552919899841_0007
19/03/18 09:21:40 INFO impl.YarnClientImpl: Submitted application application_1552919899841_0007
19/03/18 09:21:40 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1552919899841_0007/
19/03/18 09:21:40 INFO mapreduce.Job: Running job: job_1552919899841_0007
19/03/18 09:21:47 INFO mapreduce.Job: Job job_1552919899841_0007 running in uber mode : false
19/03/18 09:21:47 INFO mapreduce.Job:  map 0% reduce 0%
19/03/18 09:21:53 INFO mapreduce.Job:  map 100% reduce 0%
19/03/18 09:21:59 INFO mapreduce.Job:  map 100% reduce 100%
19/03/18 09:22:00 INFO mapreduce.Job: Job job_1552919899841_0007 completed successfully
19/03/18 09:22:00 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=56
                FILE: Number of bytes written=287589
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=30691
                HDFS: Number of bytes written=39
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=3601
                Total time spent by all reduces in occupied slots (ms)=3474
                Total time spent by all map tasks (ms)=3601
                Total time spent by all reduce tasks (ms)=3474
                Total vcore-milliseconds taken by all map tasks=3601
                Total vcore-milliseconds taken by all reduce tasks=3474
                Total megabyte-milliseconds taken by all map tasks=3687424
                Total megabyte-milliseconds taken by all reduce tasks=3557376
        Map-Reduce Framework
                Map input records=223
                Map output records=6834
                Map output bytes=100232
                Map output materialized bytes=56
                Input split bytes=151
                Combine input records=6834
                Combine output records=3
                Reduce input groups=3
                Reduce shuffle bytes=56
                Reduce input records=3
                Reduce output records=3
                Spilled Records=6
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=473
                CPU time spent (ms)=2760
                Physical memory (bytes) snapshot=642076672
                Virtual memory (bytes) snapshot=3164168192
                Total committed heap usage (bytes)=546832384
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=30540
        File Output Format Counters
                Bytes Written=39
The standard deviation is: 27.04312972098201
[cloudera@quickstart ~]$
```